W Y D A W N I C T W O P O L I T E C H N I K I Ś L Ą S K I E J W G L I W I C A C H

ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2018
Seria: ORGANIZACJA I ZARZĄDZANIE z. 122

ETHICAL CASE FOR SELF-DEFENSE ROBOTS

Maciej ZAJĄC

University of Warsaw, Department of Ethics, Institute of Philosophy; maciekzajac1@gmail.com

Abstract: Advances in robotics and Artificial Intelligence (AI) make utilizing robots in domestic law enforcement and self-defense increasingly viable. As with any set of technological developments, this opens up a space of possibilities for their licit and illicit use and creates a need for ethically-informed, effective and enforceable regulation. What is unique is the technology's potential to solve a persistent issue necessarily impacting all self-defense and law enforcement executed by humans – the License to Overkill and the Response Time Trilemma. The paper starts by explaining what these issues are and how they thwart our attempts to conform with an ethical ideal of self-defense. It then outlines the particulars of the technological solution and describe its great promise. Finally, possible objections and the regulatory measures that would limit the negative impact of self-defense robots while preserving the social and ethical benefits of their sensible employment are discussed.

Keywords: violence, crime, law, law enforcement, self-defense robot, robotic autonomy, artificial intelligence, technology, ethics.

License to Overkill

The following is, I believe, a correct account of moral principles underlying the right to self-defense and its exercise: 1) The Attacker forfeits his rights vis-a-vis Victim(s) or Vindicator(s)[1] to a degree that is necessary to render him harmless. 2) While Attacker also forfeits his rights to a degree necessary for him receiving a just and fair punishment, he does not, for due process reasons, forfeit those rights vis-a-vis Victim(s) and Vindicator(s), and thus lynching and other forms of summary punishment always violate his rights, especially as 3) the Attacker always retains certain rights, including but not limited to procedural ones[2]. 4) The Attacker's immoral actions do not have the power to burden the Victim(s) or Vindicator(s) with any additional moral duties towards the Attacker. 5) It is consequently just and fair to burden the Attacker with

[1] Any person engaging in legitimate other-defense.

[2] Even the supporters of capital punishment, who believe that to commit certain violent crimes is to forfeit one's right to life, do not believe a death row convict to have forfeited his right not to be tortured, or not to be raped.

costs and/or risks of preventing, mitigating or undoing the wrongs streaming from his actions – provided that such costs and/or risks are broadly proportional to the amount of actual or potential damage involved. 6) It is thus morally preferable that Victim(s) or Vindicator(s) deal the attacker excessive harm rather than allow themselves to be harmed by him (again, within limits of broadly-understood proportionality). 7) Calibrating the amount of force so that it would render the attacker harmless without doing him any excessive harm is impossible or extremely hard for Victim(s) and Vindicator(s). Consequently, it follows from #6 & 7 that 8) Victim(s) or Vindicator(s) are allowed and are to be expected to deal excessive harm to the Attacker in the process of defending themselves, and are fully excused for it as long as such harm is casually instrumental in stopping the Attacker, broadly proportional to the danger he was posing or might have been assumed to pose in the epistemic circumstances in which the decision was made, and not demonstrably wanton. I shall dub 8) the License to Overkill.

#1, 2 and 3 are basic, commonsensical principles binding in every society that has a functioning apparatus of justice[3], discussed and endorsed exhaustively in the literature (Coons, & Weber, 2016). #4 is inherent in the fact that we do not have the ability to successfully claim benefits from others at will, let alone by harming them. #5 is really an extension of #4 – if the Attacker cannot burden others with the cost and risks streaming from his wrongs, it follows he needs to accept it himself – hence #6. #7 is what counts as a non-ideal principle, one resulting from contingent features of the world – such as details of human biology or the speed with which the police can be expected to travel – rather than from more universal and timeless considerations (Pattison, 2016) (to an extent that my account differs from standard ones, it is so because of the ethical importance I assign to the considerations of non-ideal theory, especially epistemic ones). As it is preferable that all the harm resulting from the cost of effective defense against his actions befell the Attacker; and as the only way to produce such an outcome is allowing the Victim to maximize her chances of coming out unharmed at the cost of harming the Attacker more extensively that might be necessary for a knowledgeable, cold-blooded and skilled defender – since no Victim should be punished for failing to display average levels of martial skill, judgment and composure, let alone such that cannot be expected even of security professionals. Acceptance all of the above begets the License to Overkill.

An example: say Grandma Tammy is suddenly awoken at 3 a.m. by a man who enters her room with an intention (known to him, but unknown to her) to merely steal her jewelry. The man would rather not wake Grandma at all. Grandma awakens, sees him three steps from her bed and within a few seconds shoots and kills him with a sawed-off shotgun she keeps under her pillow. That is certainly an overkill – we do not view death as just punishment for theft, and the man loosing his life over a couple of old-fashioned necklaces is clearly a tragedy. If the circumstances were different – if auctioning off the jewels was the only way to provide this

[3] I.e., one able to justly and competently take over from the Victim and Vindicators when its representatives arrive at the scene.

man with life-saving medication – we would urge grandma Tammy to donate them to this end, perhaps even believe she is morally obliged to do so. Yet she could not have been certain that this was a mere robbery and not a murder attempt, and even if she knew 99% of such cases are in fact robberies, and could reason on this fact with clear mind at 3 a.m. (not a standard we can hold her to), she should not be required to run a 1% risk of dying for the sake of a man who forced her into this terrible dilemma while trying to harm her. Grandma has a reason to fear for her life, she is innocent of this reason having come to be (while the man is guilty of putting her in this tragic position), and her only effective way of rendering herself safe is using the shotgun. Thus, she gains a License to Overkill.

Grandma Tammy's case is, admittedly, located on the far end of the self-defense spectrum – but it still shares general characteristics with many, perhaps even with most self-defense cases. As most victims, she is significantly handicapped in her fight against the Attacker (that is exactly why she has been chosen to be victimized); she has very limited access to information about the Attacker, his abilities, resources and intentions; and she needs to process that limited amount of information extremely fast and under very significant stress. The severity of these limitations does vary from case to case, but generally allows the Victim an excuse for actions we would never condone were they to be retributive, rather than defensive/risk-mitigating, in their nature.

There still remains the issue of Grandma Tammy owning a shotgun. Given the physical disparity between a usual perpetrator – a man close to his physical prime – and the most vulnerable members of our society, such as the elderly or women, a License to Overkill is never truly granted without providing potential Victims with reasonably easy access to weapons they might be expected to use effectively despite their physical limitations. This is when things get really messy, especially from the consequentialist perspective. I have so far discussed only the apportionment of risk between the Victims, the Vindicators and the Attackers, but when the right to bear arms comes into play, the risk is distributed much more broadly and unpredictably among all members of the society, toddlers included. Mass-shootings, gun-handling accidents, spur-of-the-moment suicides – we need to take the damage created by all these into account[4] and if we do, there is a case to be made that an average person would be safer without universal access to firearms[5] (especially if such access would be granted to people without proper training, psychological capacities and moral qualifications). Passing judgment on these difficult issues is beyond the scope of this article, but we need to stress that they make granting the License to Overkill a very difficult choice.

[4] According to US Center for Disease Control and Prevention, in 2015 alone the US authorities registered thirteen thousands firearms homicides and twenty two thousands suicides by firearm – www.cdc.gov/nchs/fastats/homicide.htm.

[5] Hugh LaFollette is notable for making such a case based on both fact and principle and following up with a non-ideal analysis of the gun control issue to render a set of viable, middle-of-the-road solutions. LaFollette, 2000, pp. 263-281.

Thus, though License to Overkill is grounded in serious moral arguments, its exercise is always a tragedy leading to a less-than-perfect outcome. Such an outcome is inevitable every time an amateur faces a situation that would be challenging to the best of the professionals – and self-defense is always challenging in this way. As a society we have responded by creating professional law enforcement whose overwhelming strength vis-a-vis a private citizen would enable a much more measured response (a team of riot police is not threatened by a looter in the way that Grandma Tammy would be, and so has no need, and no right, to use sharp ammunition while subduing him). Given the cost of raising, training and sustaining a professional police force, it is necessarily limited in number, and consequently absent in the first crucial moments, minutes or even hours of a violent attack. This gives rise to a second, broader problem with contemporary self-defense/law enforcement paradigm:

The Response Time Trilemma

No contemporary society can avoid one of the following:

A) Granting the Victim and/or Vindicators a right to effective self-defense – entailing the License to Overkill and a right to bear at least some kinds of lethal weapons.
B) Leaving a Victim at the mercy of the Attacker within the time period it takes law enforcement to reach the scene.
C) Making the response time negligibly short by making law enforcement agents ubiquitous[6].

I have already discussed the flaws of option A. Option B entails its very opposite – a state-enforced denial of the Victim's right to effective self-defense. To do so, a legal system does not need to take from the Victim every single defense prerogative she has – stripping her of a single critical prerogative, such as a right to own a firearm, will suffice in many cases (like the case of Grandma Tammy). Indeed, the most common restrictions of this kind are limitations or bans on possession and bearing of arms by citizens. It may be plausibly argued, based on empirical data and non-ideal theory principles, that a well-crafted system of such bans and restrictions is in fact the best possible solution – but whatever the strength of such an argument one must be clear about the trade off involved. Denying law-abiding citizens reliable access to firearms simply means that the most vulnerable need to count on either never becoming Victims, or being rescued by the police quickly enough.

There are other ways of limiting the Victim's self-defense powers. A legal system may insist on strict proportionality of either means or effects of violence. Banning the Victim from defending herself with a knife from an Attacker kicking and punching her, even if the Attacker

[6] While not giving the phenomenon a name or discussing it explicitly, Jeff McMahan recognizes the game-theoretical forces behind while making a utilitarian case for option B – McMahan, 2015.

is twice her size and strength, would be an example of the former; banning her from punching a person who merely slapped her repeatedly would constitute the latter. Yet another way is forcing a Victim to be strictly reactive throughout her encounter with the Attacker – to always shoot second, literally. This in turn is just an extreme example of placing epistemic burdens upon the Victim – which in its less extreme versions requiring her to prove the Attacker had certain intentions, and was able to carry them out, and perhaps that she tried to actively avoid the confrontation. Last but not least, the ability of Vindicators to come to the Victims rescue without fearing legal repercussions for actions that would be accepted or even praised had they been performed by law enforcement agents may be severely restricted.

From a purely consequentialist point of view, the types of regulations outlined above may not appear very problematic. After all, in the XXI century violence is rare in all those societies that do not experience significant problems with governance or the rule of law – such as Western or Far Eastern democracies. If one wants to protect herself or her family from death or injury, the best way to go is eliminating processed sugars and cigarettes from their lives, making them fasten their seat belts and wear their bike helmets, and safeguarding them from opioid or alcohol addiction. Violent crime is orders of magnitude less likely to cause a persons death than cardiovascular disease or cancer, and about five times less likely than a car accident is[7].

Still, this perspective ignores both facts and sensibilities that need to be accounted for. Firstly, the average level of safety does not hold in many places and for many vulnerable groups. For example, the scale of violence directed at women is still coming to be accurately represented in the statistics and the public consciousness[8]. The most methodologically robust sources – wide scope, random respondent sample victimization surveys – show more than 1 in 5 EU women have reported experiencing violence at the hands of an intimate partner, while 1 in 20 report having been a victim of rape[9]. These are prevalence levels that make it perfectly rational and justified for an at-risk person to seek effective means of self-defense. When several factors increasing vulnerability combine – as in a case of a woman of low socioeconomic status living in a neighborhood with sparse police presence – a person can ran an even more significant risk of becoming a victim of violent crime (including sexual assault) during her lifetime. In such a

[7] According to Eurostat, in 2014 the average EU death rate from heart and circulatory disease amounted to 500 annual deaths per 100 000 people. Combined cancer death rate was 350 annual death per 100 000 people. After being halved in a decade, the figure for transport accidents stood at 6 in 100 000, and that for suicides at 11. In comparison, deaths caused by violent offenses amounted to 0.7 annual deaths per 100 000 people (albeit it covered only the deaths registered by the police). Exact data for particular member countries available at Eurostat website: http://ec.europa.eu/eurostat/statistics-explained/index.php/Causes_of_death_statistics#Publications, http://ec.europa.eu/eurostat/statistics-explained/index.php/Crime_statistics.

[8] Women much more frequently become victims of sexual and domestic violence, crimes that go disproportionally under-reported and under-registered due to both objective methodological challenges and systemic indifference and/or hostility towards victims. As stated by the EU Agency for Fundamental Rights report on the issue, "official crime statistics say more about official data collection mechanisms and the culture of reporting rape than they do about the 'real' extent of rape" – Von Hofer, 2000, pp. 77-89; Yung; EU Agency…, 2014, p. 13.

[9] EU Agency…, 2014, pp. 20-22.

situation a potential victim has every right to aim at decreasing such a risk by all means at her disposal.

One must also understand that human perpetrated evil is for many if not most people a more terrifying prospect than naturally occurring health problems or accidental harm, being intentional and so an assault on the Victim's very dignity and sense of self-worth and social belonging[10]. For this reason that the threat of being raped is simply incomparable to a threat of suffering a car accident. Those especially sensitive to the prospect of violent crime (especially former Victims and the friends and family members who witnessed the damage inflicted first-hand) may need the means of defense against it to attain a level of psychological security needed for normal functioning. Such cases being relatively rare in most developed societies may lead to those needs being outweighed by other considerations; yet we may not pretend they are not part of the picture, an aspect that may by omitted in the ethical analyses of the issue.

While it may be fashionable to discard doubts about moral and practical viability of option B, option C never lacks detractors – and for good reasons. The levels of state control and paramilitarization of daily life it would require have historically been economically debilitating, damaging of the social fabric, and conducive to tyranny, either in the form of a centrally-governed police state or of the over-sized law enforcement degeneration into a network of independent, parasitic local structures being themselves the principle source of violence and human rights abuses.

Given how problematic – morally, politically, economically, organizationally – option C is, it has rarely been seriously entertained as an anti-crime measure. Yet the ascent of cheap self-defense robots may change this.

In summary: when creating a system of domestic security and law enforcement, every society faces a tragic decision that I call Response Time Trilemma: it has to choose either to establish virtually ubiquitous police presence; to empower the citizens to effectively defend themselves against violent crime in the time it takes law enforcement to arrive at the crime scene (which necessarily involves, as I have discussed, giving them access to weapons and a license to defend themselves with means that stretch the limits of proportionality); or to leave the citizens to a large extent defenseless until the arrival of the police. I have discussed the tragic nature of those trade offs, arriving at the conclusion that at present every possible system of domestic enforcement will necessarily lead to one of those troubling capitulations to the practical realities of self-defense and law enforcement. I will now proceed to argue that such capitulations will no longer be necessary – and therefore no longer morally allowed – for societies that have robotic security platforms at their disposal.

[10] The Victim's nearest and dearest, though free of physical harm, frequently suffer great if not comparable levels of psychological and moral injury.

Robotic Revolution in Self-Defense

Speaking of Robotic Revolution in self-defense, one does not need to entail the use of lethal and/or autonomous robots by ordinary citizens – although such a prospect is no longer a mere fantasy. The robotic self-defense weapons and accessories may be safely expected to occupy a long and dense spectrum of combat ability and lethality, with a large swath of this spectrum consisting of devices incapable of life-threatening action against humans. This section will discuss the shape that space of technological possibilities will likely take.

The spectrum starts with software systems connected to a network of sensors and capable of automatically calling the police, blocking the Attackers way of access/escape, monitoring the incident and following the escaping attacker while transmitting the footage to the law enforcement. A drone capable of shooting video of a robbery and following the robber on his escape route for a few kilometers is already feasible technologically, since all this requires is joining existing capacities of small hovercraft with object recognition software. While such platforms would do little to assist a Victim directly, their prevalent use would significantly raise the probability of perpetrators getting caught and thus deter criminals susceptible to rational considerations. They could also substantially increase the speed, quality and certainty of law enforcement response.

Next on the spectrum would be devices increasing the effectiveness, safety and ease of use of both lethal and non-lethal weapons. Ballistic calculators fitted onto intelligent scopes have been proven to greatly improve shooting performance regardless of experience level (Marks, 2013); sophisticated augmented reality systems could make using guns, tasers or even melee weapons much easier and more intuitive for untrained amateurs. That would not only give the Victims (and law enforcement officers) a greater fighting chance; it would also decrease the extent to which the Victims need (and so are entitled to) their License to Overkill.

The trend his clear – the more these technologies matures, the more precise, faster and more intelligent robotic devices of any kind get – the less lethal and harmful they need to be to subdue an Attacker. That is especially true of crimes committed in a state of intoxication, emotional upset or temporary insanity, when the perpetrator does not plan in advance, and so comes to a fight armed lightly or not at all. In these cases even relatively unsophisticated robotic weapons may give the victim a decisive advantage without relaying on any kind of overkill.

Yet another category consists of stationary weapons with limited autonomy. Things like stationary gun turrets capable of target identification and acquisition are already employed by Korean and Israeli armed forces (Velez-Green, 2015); while such devices are ethically controversial, there non-lethal versions, armed with taser-like weapons or water cannons, need not be so. One may argue that humanity has uncontroversially employed area-defense, borderline-lethal weapons for thousands of years now in the form of guard dogs. Meeting a

defense robot armed with a net or even a taser, while certainly unpleasant, seems preferable to meeting a rottweiler – and she who may do more, may also do less.

Non-lethal stationary obstacles, surveillance bots and AI-driven human capacity enhancers such as intelligence scopes may emerge and be put into actual use much faster than fully-fledged self-defense robots equaling or surpassing human capacity for movement and opponent incapacitation, being both easier to design and more in line with existing self-defense laws giving Victims more self-defense powers on their own private premises.

Security robots with full capability for movement in two or three dimensions coupled with potential for identifying attackers and effectively engaging them with one or more weapon types would top the defense bots' capacity spectrum – and so remain the farthest from technological viability. We may be far from fielding those even in the military realm – perhaps even decades away[11]. But given that all the component technologies needed to make such platforms a technological possibility are multi-use and necessary for developing a number of key civilian technologies of tremendous importance, we will eventually be faced with a choice to either suppress these kind of weapons or incorporate them into existing legal and ethical framework. It is therefore advisable to examine their potential to blunt each horn of the Response Time Trilemma.

Response Time Trilemma Dis-Horned

Let us start with a Right to Effective Self-defense. As already shown, robotic technologies may empower the Victims, allow them to be much more precise in using force and, by decreasing the risk posed by Attackers, also decrease the need, and warrant, for the more extreme self-defense measures. Still, two powerful arguments can be offered against the claim that this would reduce the amount of violence and harm inflicted onto the Victims – arguments that are at least partially successful in presently existing technological context. Yet I will argue that the analogy does not carry over to robotic defense systems.

First, making a class of weapons available to the general public entails making it available to criminals. While access restrictions may be put into place, the most dangerous kinds of criminals will always find a way to acquire these weapons if they are around in large enough number. Even if my home is guarded by a defense robot, what chance will it have against a dozen robots brought into a fight by professional outlaws?

That argument, while initially plausible, cannot go as far as it does in relation to firearms. A confrontation between men armed with assault rifles will, on average, result in a vastly more

[11] Though perhaps not that far – general Mick Ryan believes advances in combat robotics capable of reducing an infantry battlegroup's manpower needs five-fold by 2030, by which date he anticipates 'thousands' of robotic systems to be employed by each such unit – Ryan, 2018.

tragic outcome than the same fight being resolved with bare fists. Thus it makes no sense to increase access to assault rifles if a large enough portion of criminals was indeed bound to acquire them. However, because defeating a robot does not require defeating the person it protects (especially if the robot is acting autonomously, rather than being remote-controlled), a robot-on-robot shoot-out has a large potential of being bloodless, with the side rendered robot-less surrendering to the opponent's will. The more an average defense robot outmatches a human in fighting capacity, the more probable this becomes. This alone is an outcome worth pursuing, yet is coupled with the bots tilting the balance in Victim's favor in situation where the attack is an effect of an unplanned outburst. In result access to self-defense robots would benefit Victims greatly, even if top-layer criminals would also make use of these platforms. Grandma Tammy is much less likely to be attacked by a mafia boss than by a petty hoodlum or her drunken husband.

The second argument against citizen access to any transformative weapon class focuses on their offensive potential. Allowing citizens to posses defense robots is equivalent to permitting them to command a ruthless, fully-obedient militia. Everybody who listens to Grandma Tammy's dinner monologues knows how anti-Catholic she can get. If she finds herself in command of a squad of machine warriors, what will stop her from going full Cromwell on the papists the very day she receives a terminal diagnosis?

The measures that ought to be taken to make robot-enabled offensive action against fellow citizens impossible are the same measures that should be used to stop criminals from employing defense robots to their own ends. The first is limiting the number of such platforms a household, institution or commercial establishment are allowed to own. While owning fifty AR-15s does not give an individual significantly more power over one's fellow men than owning one such rifle, owning a defense robot swarm may enable a single individual to inflict hundreds of casualties. Even more important is limiting these platforms' area of operations to within the limits of their owner's home or business. This may be done by equipping each bot with a kill switch that will automatically turn it off if it leaves such an area – a solution enabled by equipping each robot with inertial positioning system, thus eliminating the potential strain on the GPS transmissions bandwidth. Both the numerical and spatial limitations should be swiftly legislated and vigorously enforced.

Such legal arrangements would grant potential Victims protection within their own homes, while allowing (and further down the line requiring) property owners to offer all their guests and patrons protection of their defense robots. Private security outside of home could be bolstered by making means of travel such as cars and motorcycles equipped with non-lethal defensive capacities and smart sensor suits. As for public space, the same types of robots could multiply police presence without a need to pull many more humans out of non-security work force, induce them into joining law enforcement, train them, and provide them with salaries, healthcare, insurance and pensions. Robot policemen would probably feature a high unit-cost, especially before the technology fully matures and becomes widespread enough that the

economies of scale can be relied on; but they will work non-stop, suffer no exhaustion, sleep deprivation, feel no fear or anger, and will be much more manageable and accountable than human cops are (for this end all units should come equipped with "black boxes" that would store records of all their activities). Such expendable defenders would also be capable of taking more risk in defending Victims of violence, since safeguarding officers lives would no longer be part of the equation.

Thus, sensible and well-guided development of defense robots could bring forward a world of reliable, effective, proportionate and non-lethal self-defense, together with shortened police response times and greater law enforcement transparency and accountability – all that without putting more offensive power into the hands of private citizens and enabling criminals to engage in much more violence. Yet it also seems to empower the state, making its presence ubiquitous and constant, its surveillance powers boundless, and its ability to control its citizens absolute. Rather then solving the Response Time Trilemma, the introduction of law enforcement robots looks likely to impale its proponents on the Trilemma's third horn. I admit that one of the potential outcomes of the Robotic Revolution is just such a dystopian future. But this outcome can be avoided, and it is the very nature of autonomous robots that makes a more promising scenario possible.

The chief concern is that possession of defense robots will give too much power either to their private owners or to the police force which employs them. But it would be so only if the robots were to be directly controlled by their human handlers, and such control is not necessary for robots' being able to complete defensive tasks (just as remote control of guard dogs is not necessary). To be a competent Vindicator the robot must only know the legal limits of its actions (not being allowed to harm third parties, nor to attack opponents that have surrendered or been rendered defenseless) and the principles to guide them (breaking up all violent interactions between humans within its area of operations). If a machine would be able to follow such rules well enough, it could not be used as a tool of aggression or oppression – since its programming would not allow for offensive action – while still acting as a neutral keeper of the peace capable of stopping violence and handing its perpetrators over to human law enforcement.

An objection could be raised that making robots capable of understanding rules at this level of generality they would require equipping them with a very broad and high-performance Artificial Intelligence – unlikely to be achieved any time soon, and most probably very dangerous to implement. Yet research conducted by Robert Arkin and his colleagues opens another, simpler way. Instead of striving for AI smarter than anything achieved up to date Arkin (whose efforts were aimed at making robots compliant with the Laws of War, but may be applied to other complex sets of behavior regulations) suggested having human specialist translate laws, rules and procedures into long lists of very simple requirements. Complex conjunctions of such negative and positive conditions for taking a specific action would translate into specific self-defense scenarios. A general rule: "Arrest anybody who fires a gun" could be translated to: "if a sound meeting characteristics of a gunshot is detected, and/or if a

citizen sends a distress signal followed by screams like ‘gun’ or ‘shooting’, approach the location of the incident. If you detect a gun-like object being held by a human, issue a warning and a call to surrender. If the gun is not placed away from the human after a warning then...” etc. The large and growing computational power available would allow defense-robots to check for complicity with hundreds of conditions in real time, while a solution known as Cloud Robotics would let the bot access thousands of solutions worked out by other robots and humans in real or simulated scenarios similar to the one it is facing, pick the most successful out of all actions permitted under the rules binding it. Cloud Robotics would allow for all bots learning from every mistake ever made, and for human-made suggestions and corrections to make a lasting impact on all future actions (Pratt, 2015, pp. 51-60). In this way the robots’ software could be very far from general AI but reliably compliant with both the law and common sense. The robot force could also be retrained faster, easier and more reliably than any human equivalent, enabling experiments with various sets of behavioral rules before finally settling on the best performing one.

Translating our self-defense regulations and arrest procedures into simple conditions will in itself be a difficult-to-complete effort for *humans*, entailing not only painstaking enumeration of things obvious to every human brain for the consumption by software devoid of our built-in intuitions, but also legal, moral and political debate on fundamentals underlying use of force in our societies. I believe the necessity to seriously and productively engage in such a debate to be one of the most enticing prospects connected with applying the Arkin scheme.

Using Defense Robots Right – A Normative Framework

Steadily increasing technological feasibility of robotic solutions in self-defense and law enforcement is an indisputable trend, driven by activities of researchers, engineers, companies and institutions outside the broadly defined defense sector – developments that constitute the phenomenon of the Robotic Revolution. While governments and defense sector companies are engaged in creating robotic weapons and defense accessories – which undoubtedly increases the rate of technological progress in that area – market incentives for research and development in robotics and AI are themselves sufficient to fuel the Revolution, and to result in proliferation and affordability of both hardware and software easily adaptable for defensive purposes (Pratt, 2015, pp. 51-60; Altmann, & Sauer, pp. 117-142). Given these conditions, blanket ban on all self-defense applications of robotic devices would not only be counterproductive, irrational and detrimental to rights and well-being of law-abiding citizens, but may also prove indefensible under the legal framework of many liberal democracies, especially in the United States (Terzian, 2013, pp. 755-796) – legal challenges to any substantial restrictions are sure to come

when robotic platforms will start to constitute a viable non-lethal alternative to firearms as means of self-defense.

Governments could still choose to prohibit ownership of defense robots by ordinary citizens. Such a move is unlikely to by itself resolve very serious homeland security issues connected with increasing ubiquity of civilian-use robots, given how easily such platforms can be converted into dangerous weapons[12]. It is much less likely that the various law enforcement agencies will themselves cease to acquire and utilize robotic devices in ever growing numbers[13]. If they do not, democratic societies will face the most profound risks and threats associated with this technology anyway. Consequently, the possibility of utilizing defense bots to reinvent the landscape of self-defense and law enforcement is not something that can be simply disregarded at a policy level; defense robots will become a tempting option for any person or entity with assets to protect. Developing such machines is a high stakes game that may go awry, but playing it most probably will not be avoided. The new paradigm may be crafted right, via a thorough effort to rethink our current model and to design machines, institutions and laws that would establish and perpetuate a much better one; it may be also left to the forces of technological drift and uncoordinated individual decisions, forces that may push us in the direction of an all powerful, omnipresent state apparatus. The choice is obvious. So is the reward for making that choice – violent crime becoming obsolete within out lifetimes.

In order to achieve this end, the following normative guidelines should be adhered to and implemented when designing, developing, producing and utilizing non-military defense robots:

I. Non-lethal armaments need to be introduced as soon as their effectiveness can equal that of firearms and other lethal weapons. Governments should fund research efforts to enable this outcome as soon as possible.
II. Defense robots should operate autonomously rather than be remotely controlled, especially after their effectiveness when using non-lethal weapons reaches a level of effectiveness sufficient for making non-lethal their primary mode of operations.
III. The bots should be equipped with carefully designed and tested software, transparent and subject to review and change in case any of the robots engages in undesirable behavior. A common standard and procedural framework for reliable software testing and licensing needs to be developed.
IV. The robots' programming needs to prohibit them from ever moving against any person who is not engaged in an act of violence. The robots are not to be used to pursue or arrest of non-violent criminals, and the surveillance footage they gather while discharging their duties should not be admissible in any court of law, nor warrant any law enforcement proceedings, unless it is directly connected to a violent crime.

[12] As stated in Zając, 2017, pp. 60-71. These problems are serious enough to worry military thinkers at operational and even strategic level –Card, 2018; Hanacek, 2018; Pinion, 2018.

[13] Only in the US, „at least 910 state and local police, sheriff, fire and EMS, and public safety agencies have acquired drones in recent years" – Gettinger, 2018.

V. Each defense robot is to operate in a strictly defined, relatively small area and fitted with a kill-switch turning them off automatically after crossing the area's boundary[14].
VI. The system of institutional and democratic checks and balances placed on national and local law enforcement agencies is to be strengthened to account for a general increase in surveillance and power projection abilities made available by technological progress.

Upholding each of the above rules entails engaging in coordinated and focused research and development efforts. These should be managed by a research agency modeled after American DARPA and IARPA, prompting commercial and academic entities to undertake all promising projects regardless of their immediate commercial viability while simultaneously promoting approaches, values and best practices in line with the above guidelines and with the public interest.

Two key components that must be present so that not only the defense robots, but the entire Robotic Revolution fulfilled its promise are reliable autonomy and a reform of public accountability architecture. Those are intertwined. Reliable autonomy is necessary to ensure that our new robotic servants will not mindlessly and ruthlessly follow the whims of various human agents, endowing them with unprecedented powers, but that they will obey and, in case of defense robots, execute the law, the beneficial mechanism of protecting universal rights and realizing our common interest[15]. New architecture of checks and balances is needed to address the unprecedented disparities in power between those in control of the technological wonders and those subject to their reach.

Conclusion

Technology alone cannot solve the ethical conundrums we face, but most conundrums require a technological component to be a part of the solution. The Response Time Trilemma and the imperfections of self-defense and law enforcement undertaken by humans that compose it are one such case. The imperfections of all possible legal frameworks, and all-too-real human tragedies behind them, call on ethically-minded policy makers, technologists and all concerned citizens to design, perfect, test and introduce promising solutions as soon as they become viable. It is not only immoral but also irresponsible and short-sighted to let the forces of technological drift settle the future of human violence and our safety, or to simply capitulate to the no-longer-inevitable *status quo*. The Robotic Revolution allows us a chance to effectively abolish violent

[14] In this way, it will be impossible to effectively use the machines against mass protests or in service of any political goal.
[15] Provided such new kind of agent came to be, they could allow for laws much closer aligned with ethical ideals, as they would no longer have to account for the human flaws of persons executing them.

crime and to do that without sacrificing our civil liberties or half the GDP. As long as we are committed to the principles of non-lethality, safe-proofed autonomy, distributed control and robust checks and balances, these machines can keep us safe while making the executive apparatus of the state more transparent and accountable. Yet this promise will never materialize without a prolonged, robust and vigorous efforts across fields such as engineering, programming, law, policy and ethics to combat both unfounded doubt and reckless techno-optimism. The sub-field of civilian defense robotics will offer ample opportunities for theoretical research and practical application for years and decades to come.

Bibliography

1. Altmann, J., & Sauer, F. (2017). Autonomous Weapon Systems and Strategic Stability. *Survival, 59:5*, 117-142.
2. Arkin R.C. (2011). *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*. Technical Report GIT-GVU-07-11. Available online www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf.
3. Arkin, R.C., Ulam, P., & Duncan, B. (2009). *An Ethical Governor for Constraining Lethal Action in an Autonomous System*. Georgia Institute of Technology. Available online www.cc.gatech.edu/ai/robot-lab/online-publications/GIT-GVU-09-02.pdf.
4. Card B.A. (2018). Terror From Above; How the Commercial UAV Revolution Threatens the US Threshold. *Air and Space Power Journal, 32, 1.*
5. Coons Ch., & Weber M. (Eds.) (2016). *The Ethics of Self Defense.* USA: Oxford University Press.
6. FRA (2014). EU Agency For Fundamental Right. *Violence against women: an EU-wide survey.* Luxembourg: Publications Office of the European Union.
7. Gettinger, D. (2018). *Public Safety Drones. An Update.* Center For The Study Of The Drone At Bard College publication. Available online http://dronecenter.bard.edu/public-safety-drones-update/.
8. Hanacek, J. (2018). *The perfect can wait; good solutions to the 'drone swarm' problem,* WarOnTheRocks.com, https://warontherocks.com/2018/08/the-perfect-can-wait-good-solutions-to-the-drone-swarm-problem/, 14.08.2018.
9. LaFollette, H. (2000). Gun Control. *Ethics, 110*, 263-281.
10. Marks, P. (2013). *'Self-aiming' Rifle Turns Novices Into Expert Snipers*, https://www.newscientist.com/article/dn23571-self-aiming-rifle-turns-novices-into-expert-snipers/, 20.05.2013.
11. McMahan, J. (2015). *A Challenge to Gun Rights.* University of Oxford "Practical Ethics" Blog, http://blog.practicalethics.ox.ac.uk/2015/04/a-challenge-to-gun-rights/, 17.04.2015.

12. Pattison, J. (2016). The Case for the Nonideal Morality of War: Beyond Revisionism versus Traditionalism in Just War Theory. *Political Theory, 25th October 2016.* DOI: journals.sagepub.com/doi/abs/10.1177/0090591716669394.
13. Pionion, D. (2018). *The Navy and Marine Corps Need to prepare for the Swarm of the Future.* WarOnTheRocks.com, https://warontherocks.com/2018/03/the-navy-and-marine-corps-must-plan-for-the-swarm-of-the-future/, 28.03.2018.
14. Pratt, G.A. (2015). Is a Cambrian Explosion Coming for Robotics? *Journal of Economic Perspectives, 29, 3*, 51-60.
15. Ryan, M. (2018). *Human-Machine Teaming For Future Ground Forces.* Washington, DC: Center for Strategic and Budgetary Assessments, https://csbaonline.org/uploads/documents/Human_Machine_Teaming_FinalFormat.pdf.
16. Terzian, D. (2013). The Right to Bear (Robotic) Arms. *Penn State Law Review, 117:3*, 755-796.
17. Velez-Green, A. (2015). The South Korean Sentry – A 'Killer Robot' To Prevent War. *Lawfare,* https://www.lawfareblog.com/foreign-policy-essay-south-korean-sentry%E2%80%94-killer-robot-prevent-war, 01.03.2015.
18. Von Hofer, H. (2000). Crime Statistics as Constructs: The Case of Swedish Rape statistics. *European Journal on Criminal Policy and Research, 8*, 77-89.
19. Yung, C.R. (2013). How to Lie with Rape Statistics: America's Hidden Rape Crisis. *Iowa Law Review, 99*, 1197-1256.
20. Zając, M. (2017). Regulating Civilian-use Drones and Robots as a Serious Homeland Security Hazard. *Transformations, 3-4(94-95)*, 60-71.